

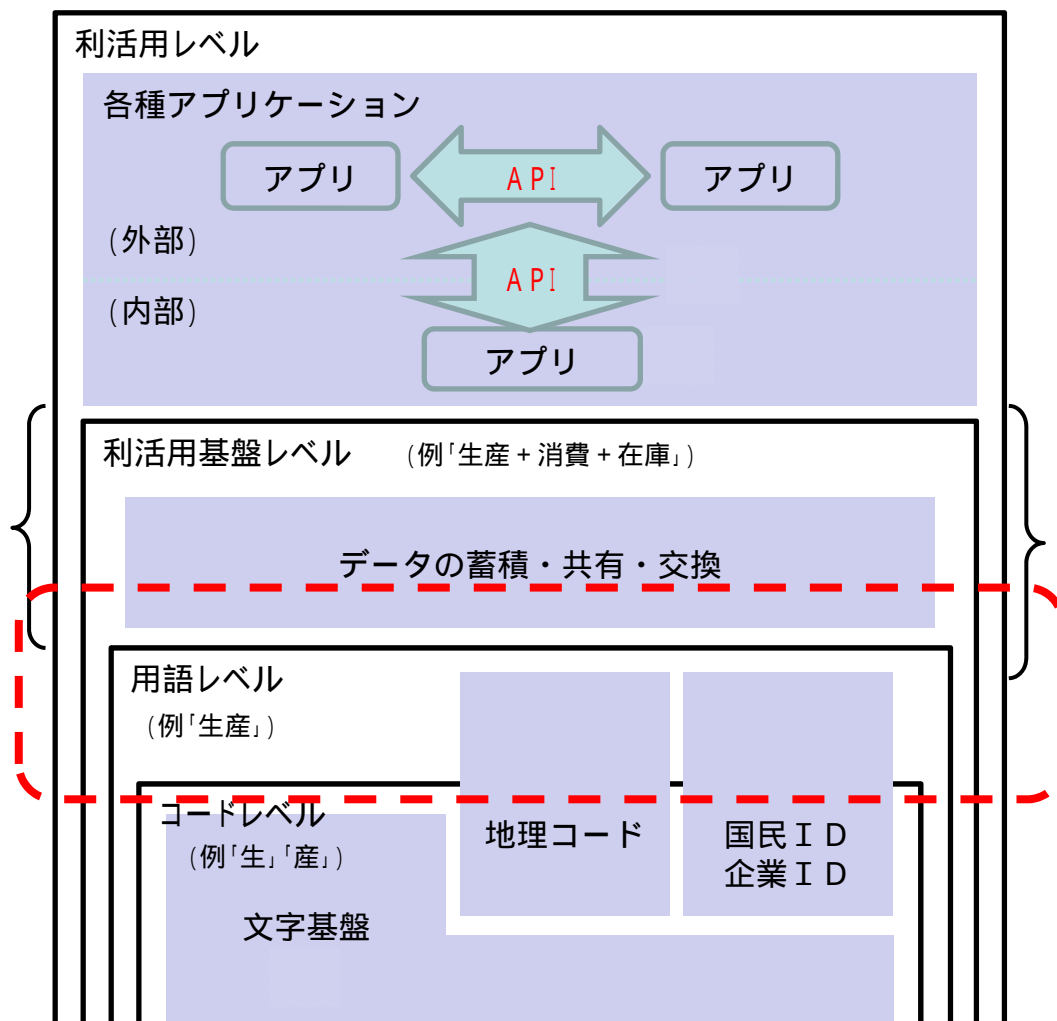
オープンデータの基盤である ボキャブラリに関する先進動向

経済産業省 CIO補佐官
平本 健二

ボキャブラリの位置づけ

- ボキャブラリは、情報処理やインタオペラビリティの中核をなす基盤であり、オープンデータ推進においても中核となっている。

➤ 米国政府のNIEM、欧州のjoinup等が強力に推進されている。
(イメージ)



利活用促進に向けた技術的な取組

- 文字基盤の統一
- ✓ IPAフォントの開発・公開 (平成23年度実施)
- ✓ 共通地理コードの検討
- ✓ マイナンバーの活用
- ✓ **公共情報交換標準スキーム、NIEM、joinup**
- ✓ 組織間での意思疎通を行うために必要となる用語レベルの意味を統一するスキーム
- NIEM (National Information Exchange Model): 米国政府が公共情報の交換に活用しているデータ交換体系
- joinup: 欧州が検討を進めている相互運用性確保のためのデータ交換体系
- ✓ **LOD (Linked Open Data) の実施・普及**
- ✓ 異なるデータ源のデータを相互に結び付け
- ✓ フォーマットの標準化
- ✓ 持続的な識別子の確保
- ✓ 公開データの機械可読化
- ✓ 機械可読なフォーマット (Excel, Word など) やオープンに利用できるフォーマット (CSV など) の活用
- ✓ 利用者向けAPIの提供
- ✓ 総務省と協力して検討を実施
- ✓ 人材の育成
- ✓ 適切なデータ公開のための加工技術とそのデータの内容が分かるデータキュレータの養成

何が変わるのか

ボキャブラリ(辞書)のない世界(従来の世界)

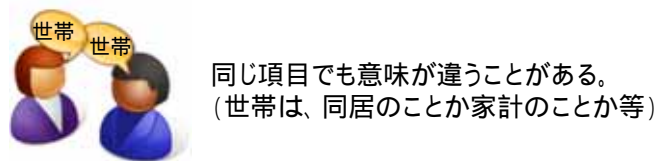
システム設計するときデータ設計が必要



データ交換をするときに形式の調整が必要



データ交換をするとき意味の調整が必要 意味のすれ違いから事故が発生



データの交換は特定者間のクローズな領域



ボキャブラリ(辞書)のある世界(これからの世界)

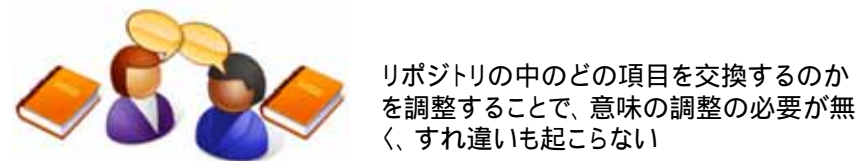
システム設計するとき基本データはリポジトリから参照



データ交換をするときにリポジトリを使うため形式の調整が容易



データ交換をするとき意味の調整が容易 意味のすれ違いが起こらない

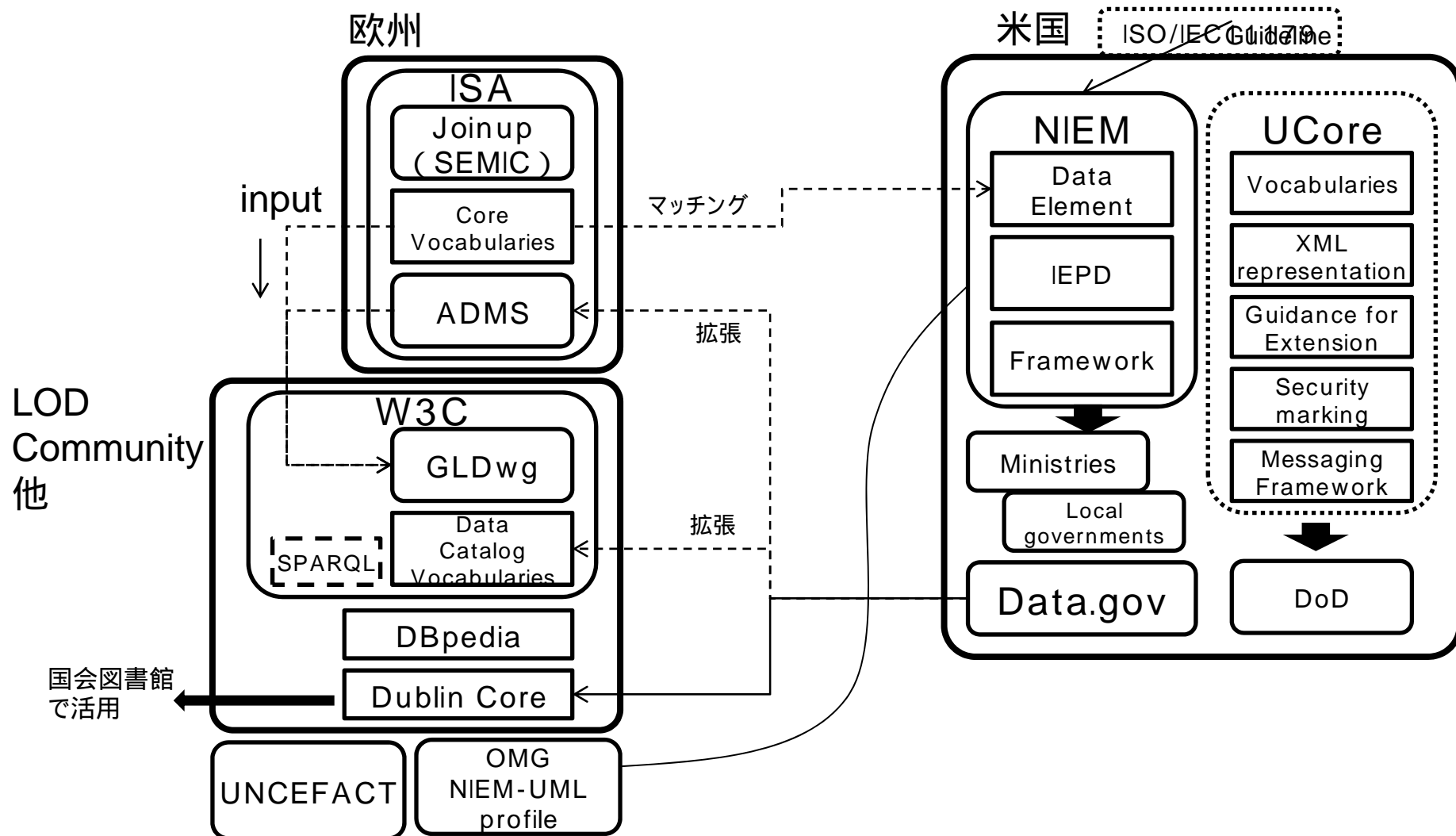


個別調整ではなくデータがオープンに流通できる



ポキャブラリを巡る世界の取り組みが進展中

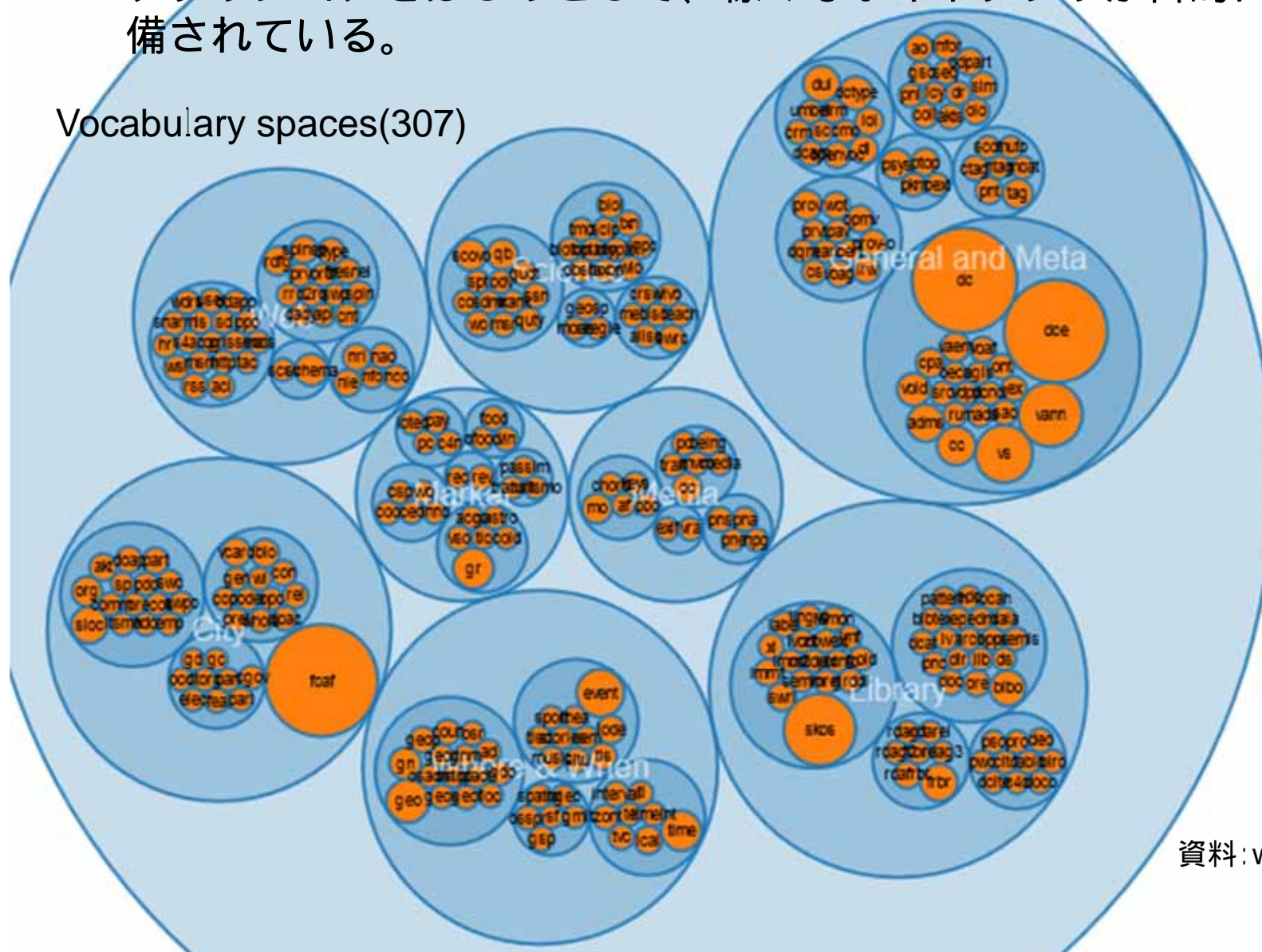
- 政府機関や標準化団体が積極的に取り組みをおこなっており、各ポキャブラリの連携も始まっている。



参考：世界の主要ボキャブラリセット

- ダブリンコアをはじめとして、様々なボキャブラリが目的に応じて整備されている。

Vocabulary spaces(307)

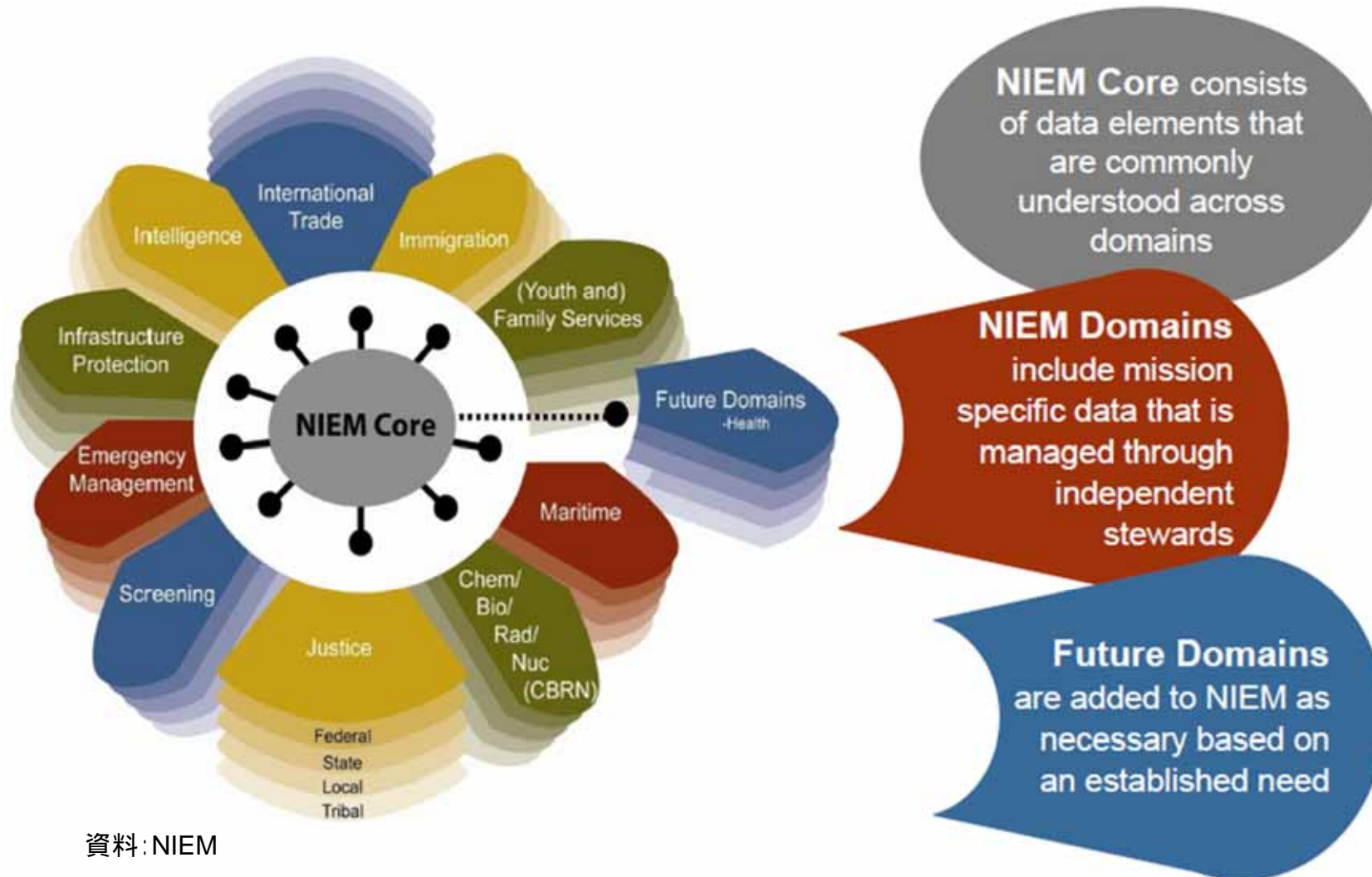


資料：www.okfn.org

米国の取り組み

NIEM (National Information Exchange Model)

- 政府内で情報交換を円滑にするためにボキャブラリの標準化を推進している。全ての分野に共通するコアなボキャブラリと分野別ボキャブラリで構成している。



資料: NIEM

- 人、場所、組織等、行政機関で使う情報と属性を構造化して管理している。

<nc:Person>
<PersonSexCode>
<PersonBirthDate>
<PersonEyeColorText>
<PersonHairColorText>
...

Has the most fields

<nc:Activity>
<ActivityDate>
<ActivityCategoryText>
<ActivityDescriptionText>
<ActivityDisposition>
...

The most often derived component

<nc:Item>
<ItemName>
<ItemDescriptionText>
<ItemConditionText>
<ItemValue>
...

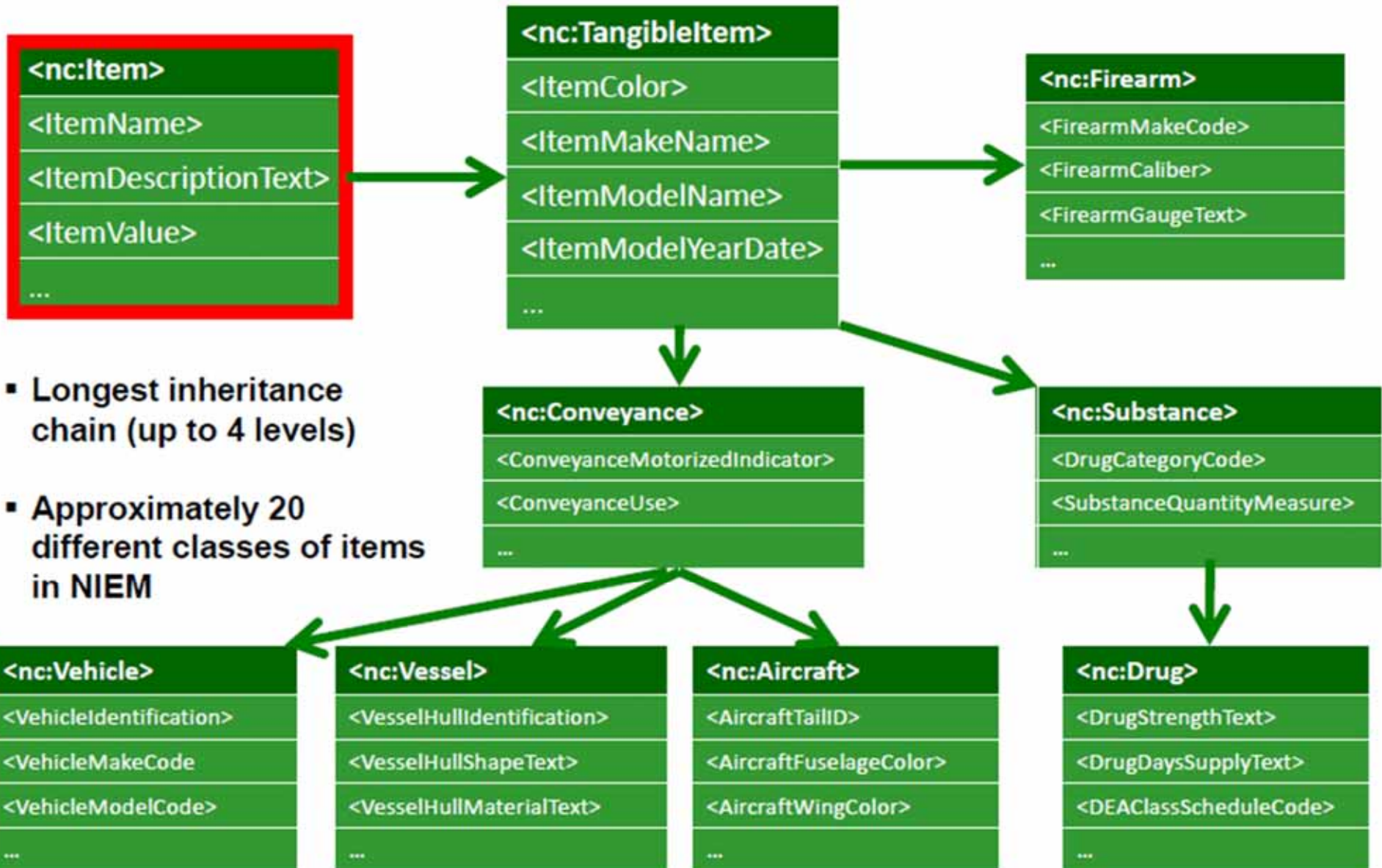
Has the deepest inheritance chain

<nc:Identification>
<IdentificationID>
<IdentificationJurisdictionText>
<IdentificationEffectiveDate>
<IdentificationStatus>
...

<nc:Location>
<LocationDescriptionText>
<LocationAddress>
<LocationCategoryCode>
<LocationName>
...

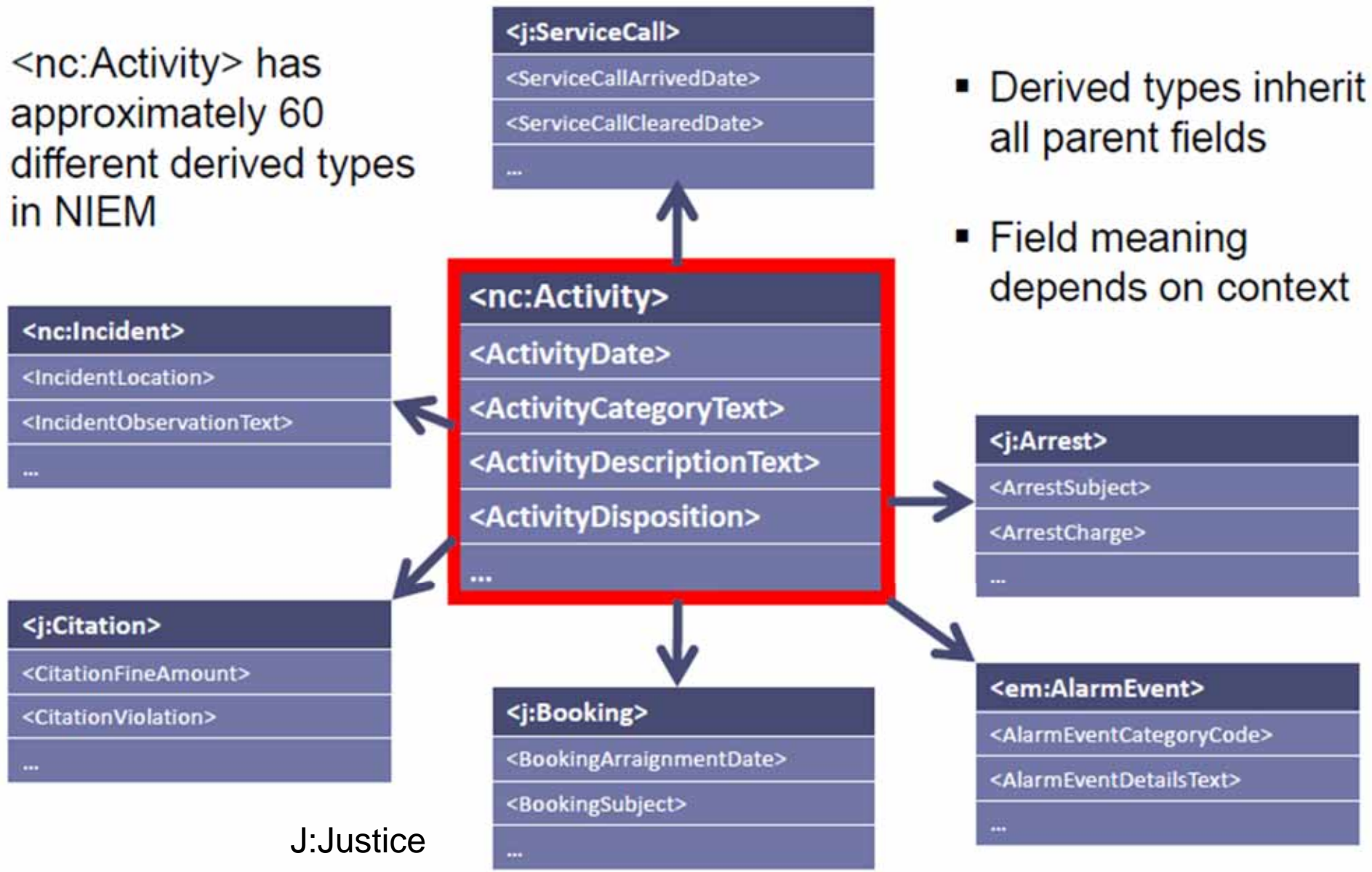
<nc:Organization>
<OrganizationName>
<OrganizationDescriptionText>
<OrganizationCategoryText>
<OrganizationEstablishedDate>
...

例 1 nc:Item



例 : nc:Activity

<nc:Activity> has approximately 60 different derived types in NIEM

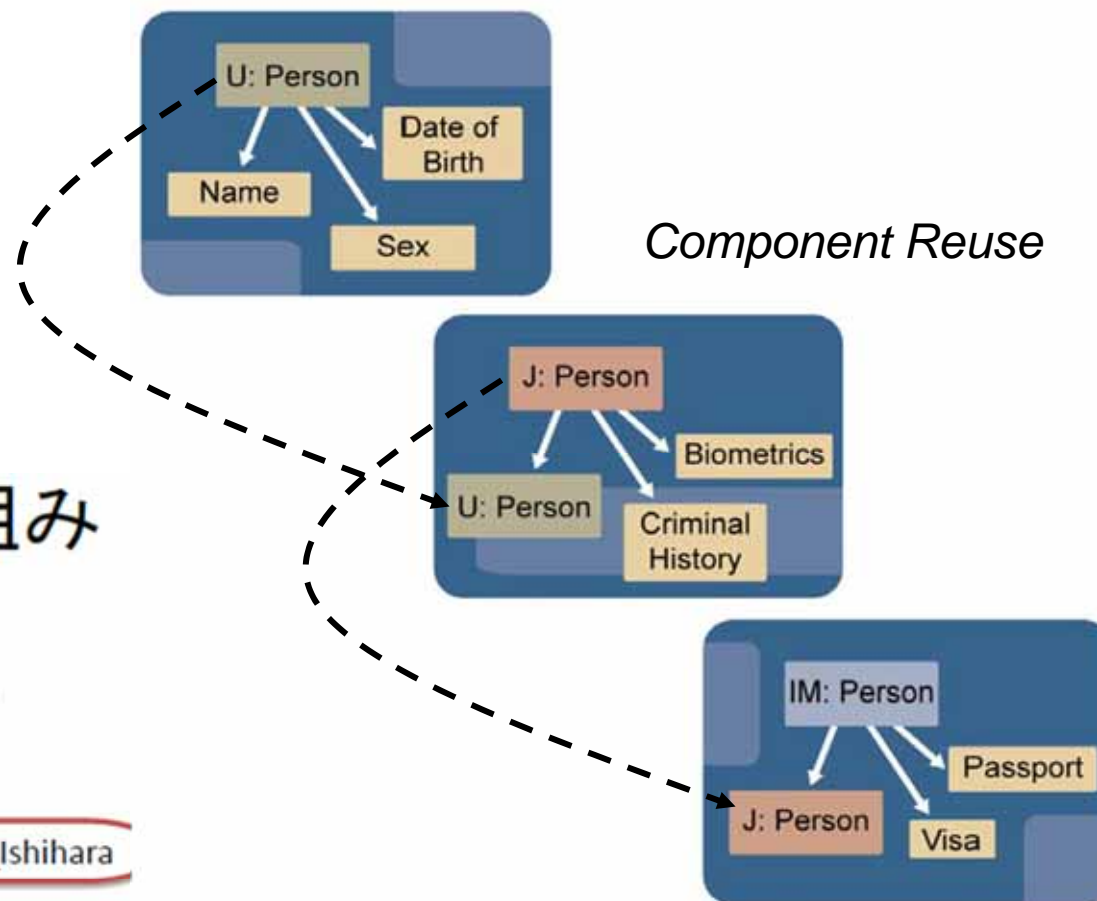


- Derived types inherit all parent fields
- Field meaning depends on context

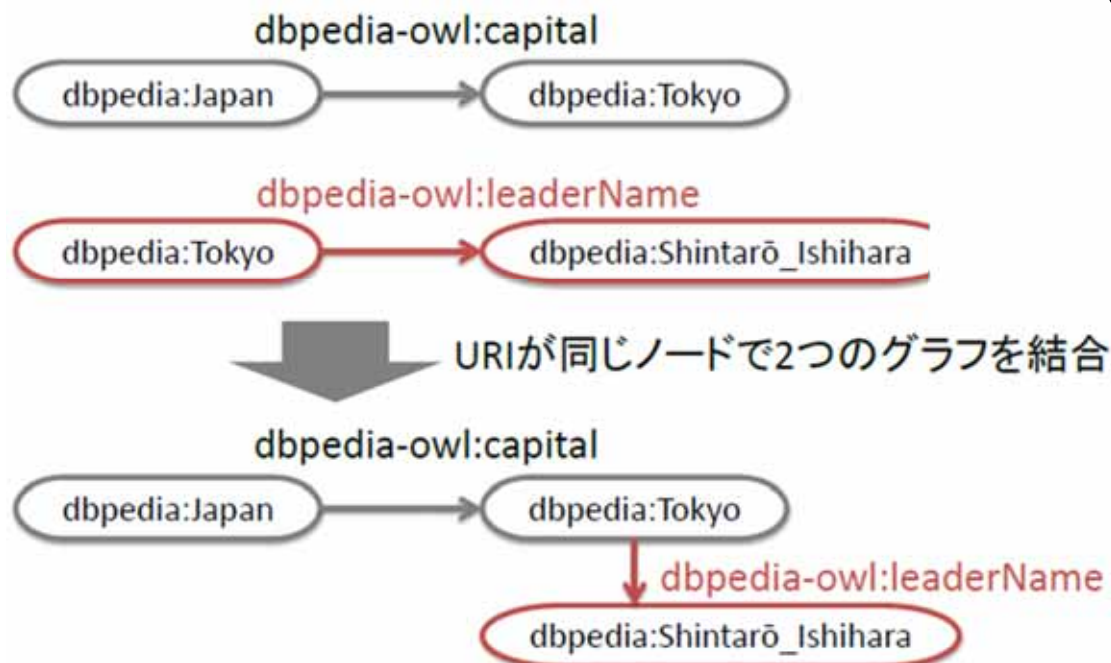
J:Justice

em:emergency

- RDFとXMLをベースにしている。
- LODと考え方は同じ
- ボキャブラリで連鎖する

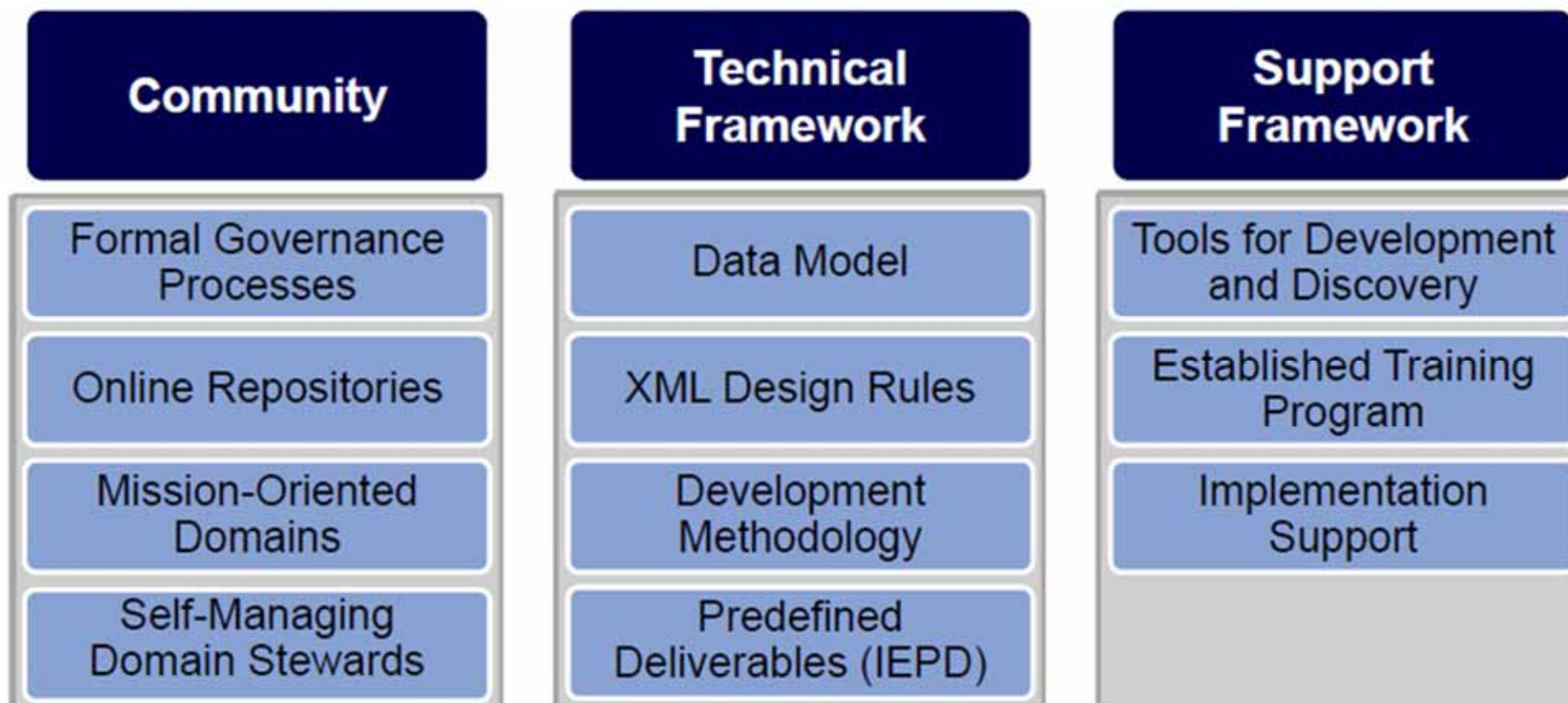


LODの基本的な仕組み



資料:「Linked Open Dataとは」
LODチャレンジ実行委員会

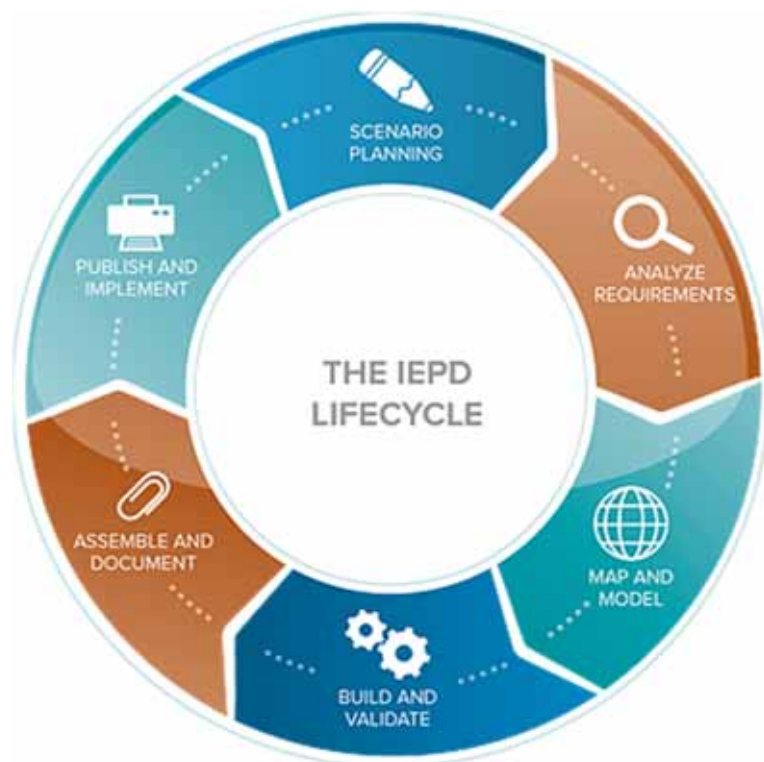
- データの定義だけでなく、導入、運用の方法、メンテナンスの方法等が体系的に整備されている。



- ボキャブラリを組み合わせ、メッセージの集積も行っている。データ交換用に蓄積されたIEPを解説するドキュメントとして、IEPD(Information Exchange package Document)を体系化している。

IEPは、データ交換用に登録されたxmlデータ

世帯主名	世帯種別	合計所得
------	------	------



IEPD 文書

NIEM 準拠の IEPD に最小限含まれていなければならない文書は、何らかの形のマスター文書と、最終バージョン以降に行われた変更を記述する変更ログです。この2つの他に、IEPD にはソフトウェア・アプリケーションに通常付属しているようなあらゆる文書を (マスター文書の一部として、あるいは別個のファイルとして) 含めることができます。以下に挙げるのはその例です。

UML モデル (シーケンス図、ユース・ケース、クラス図)

CMT

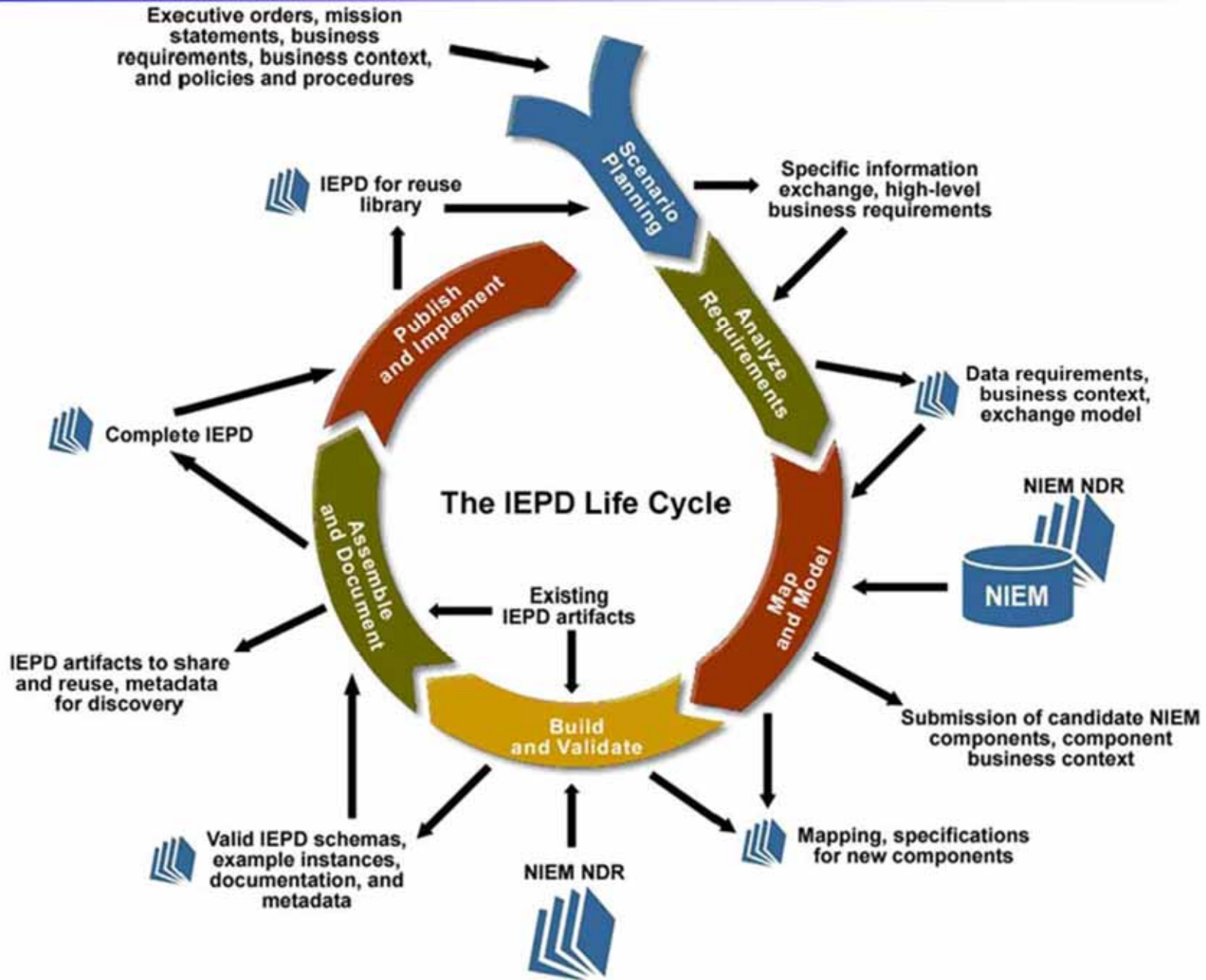
ビジネス・ルール (スキーマやモデルには表現されない、データに関する制約)

要件の定義

テストや適合性に関する宣言書

覚書および推薦状

IEPDライフサイクル



米国内での導入状況

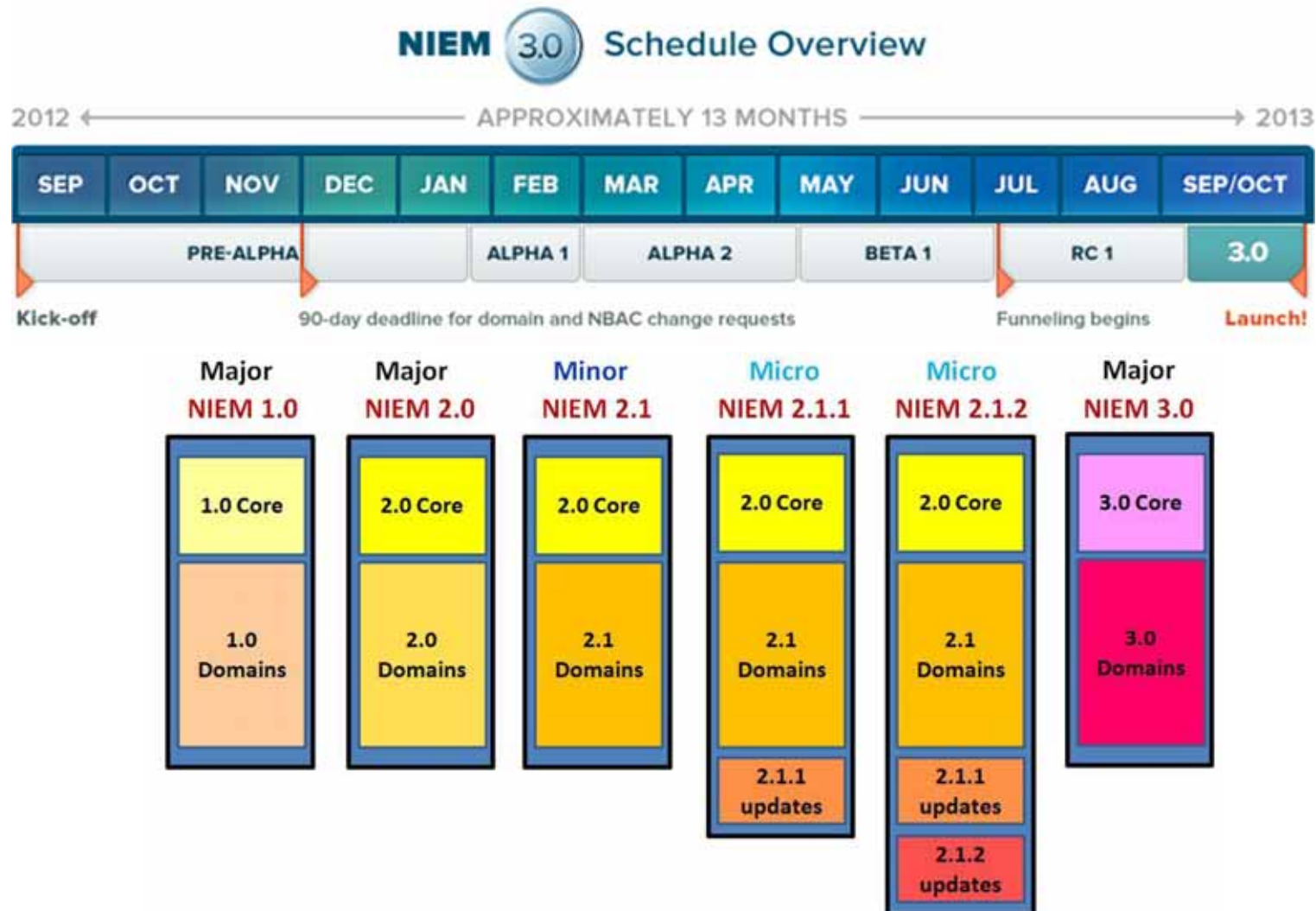
- 多くの政府機関で取り組みが行われ、地方自治体にも利用が広がっている。

Agency	Use of NIEM	
Department of Agriculture	Committed to Use	Green
Department of Defense	Further Evaluation Required	Yellow
Department of Education	Will Not Use	Red
Department of Energy	Committed to Use	Green
Department of Health and Human Services	Committed to Use	Green
Department of Homeland Security	Committed to Use	Green
Department of Housing and Urban Development	Committed to Use	Green
Department of Justice	Committed to Use	Green
Department of Labor	Committed to Use	Green
Department of State	Further Evaluation Required	Yellow
Department of the Interior	Committed to Use	Green
Department of the Treasury	Committed to Use	Green
Department of Transportation	Committed to Use	Green
Department of Veterans Affairs	Further Evaluation Required	Yellow
Environmental Protection Agency	Further Evaluation Required	Yellow
General Services Administration	Committed to Use	Green
National Aeronautics and Space Administration	Further Evaluation Required	Yellow
National Archives and Records Administration	Will Not Use	Red
National Science Foundation	Committed to Use	Green
Nuclear Regulatory Commission	Will Not Use	Red
Office of the Director of National Intelligence	Further Evaluation Required	Yellow
Social Security Administration	Further Evaluation Required	Yellow
Geospatial Line of Business	Will Not Use	Red
Grants Management Line of Business	Further Evaluation Required	Yellow
Financial Management Line of Business	Will Not Use	Red
Human Resources Line of Business	Will Not Use	Red

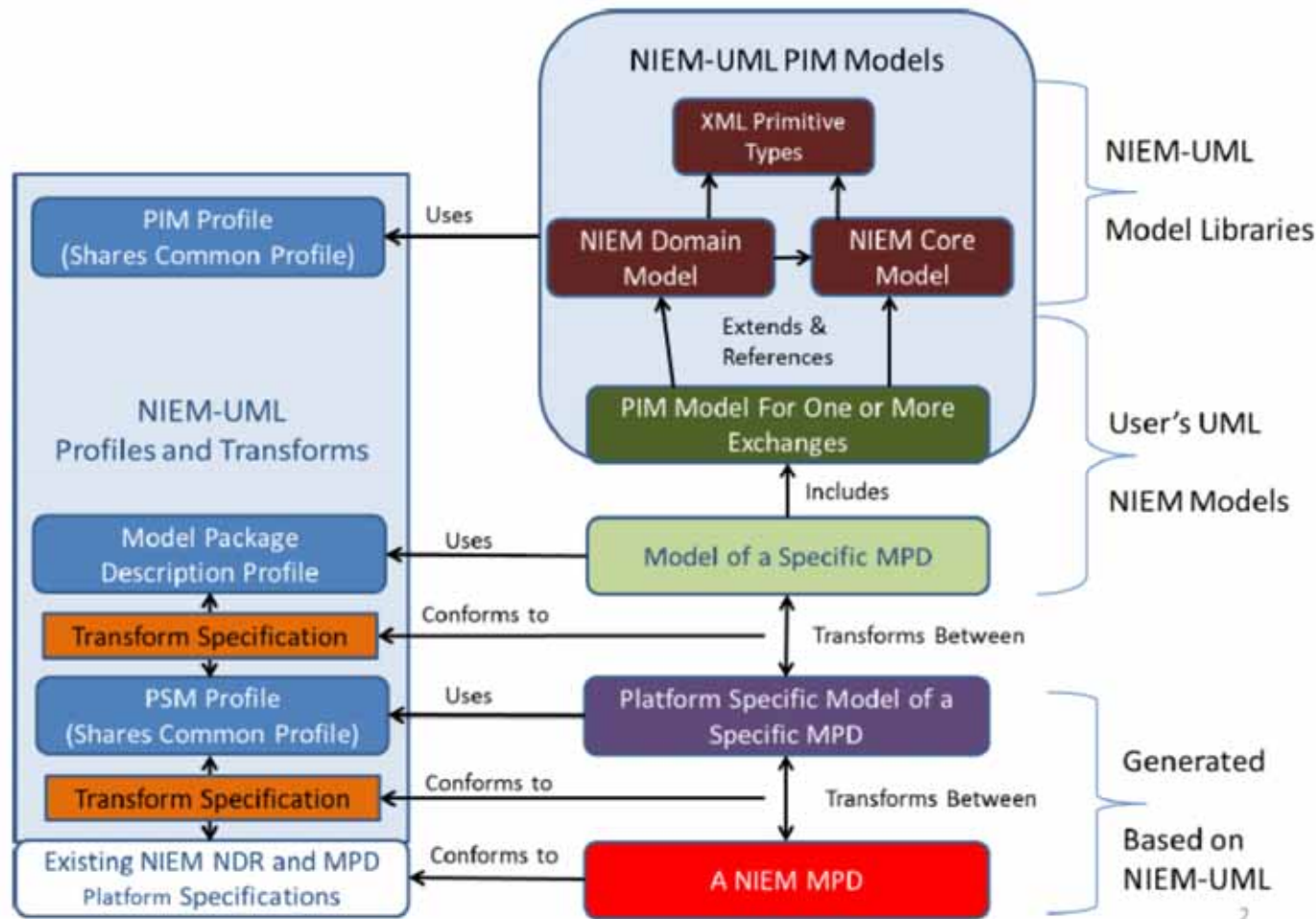
Committed to Use - ■ Further Evaluation Required - ■ Will Not Use - ■

2013秋に向けてNIEM3.0を開発中

- 継続的改善を図るため、全体のアップデートの仕組みも体系化されており、現在は、バージョン3に向けた作業が進められている。



導入を加速するUMLプロファイル



- The NIEM Platform Independent Model (PIM) Profile provides stereotypes that enable NIEM business modelers to model an information exchange in a technology agnostic way and create a NIEM PIM.
- The NIEM Platform Specific Model (PSM) Profile provides stereotypes that enable NIEM technical modelers – or, more precisely, NIEM schema modelers – to model the technical aspect of an information exchange represented in a NIEM PSM.
- The NIEM Common Profile, leveraged by both the PIM and PSM profiles, which contains the core stereotypes used to represent NIEM structures in UML.
- The Model Package Description (MPD) Profile provides stereotypes for modeling NIEM MPDs, which are the final artifacts representing a NIEM information exchange, based on either a PIM or PSM model.

欧州の動向

ISA (Interoperability Solutions for Public Administration)

■ 目的

- ヨーロッパ内の行政機関の間で、国境や組織を横断した情報交換や協調作業を効率的かつ効果的に行うことを目的とする。

■ 概要

- ISAプログラム全体で164.1 Million Euroを使って様々なプロジェクトを実施。
- 主要な柱は、「Trusted information exchange」「Interoperability architecture」「Assessment of the ICT implications of new EU legislation」「Accompanying measures」。
- 「Trusted information exchange」の中で意味論レベルでの相互運用性を検討。
(Improving semantic interoperability in European eGovernment systems (Action 1.1))
 - オープンガバメントデータ、Idや調達情報の相互運用性もTrusted information exchangeの中で検討



■ 基本理念



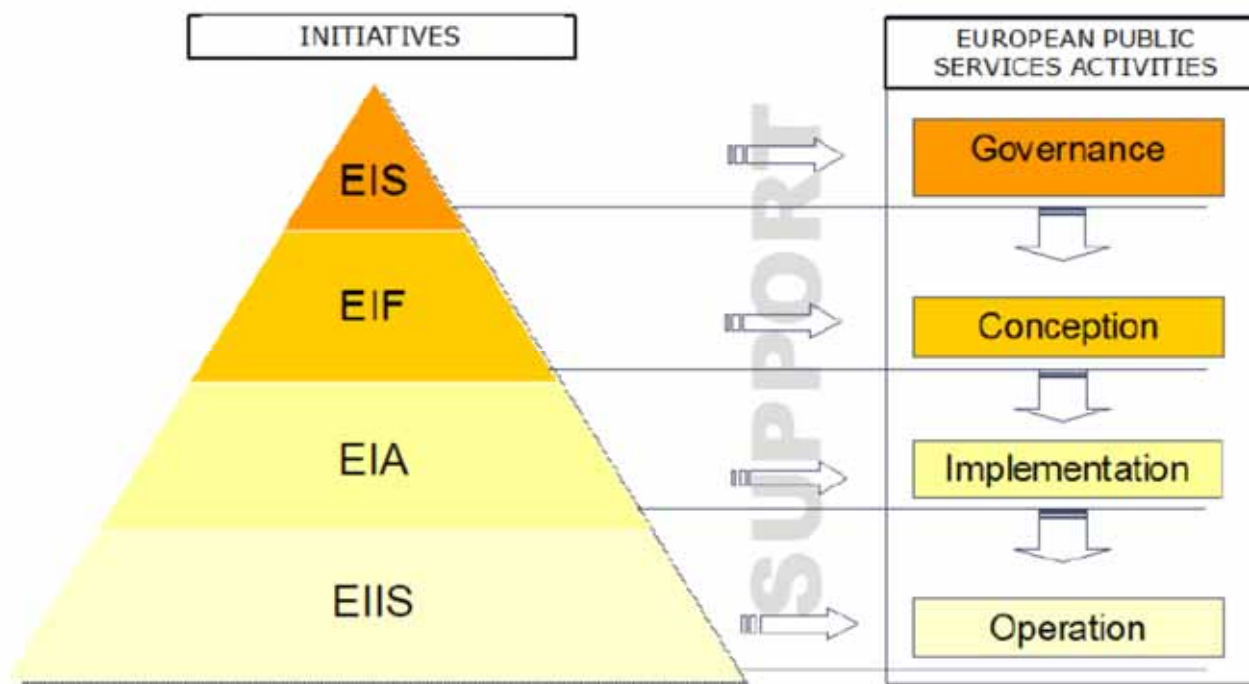
■ 主要マイルストーン

- 2012.5.12 コア・ボキャブラリを公開。
- 今後、上記をコア・ボキャブラリをW3Cの政府リンクド・データwgにインプットしていく予定

■ 取り組み状況

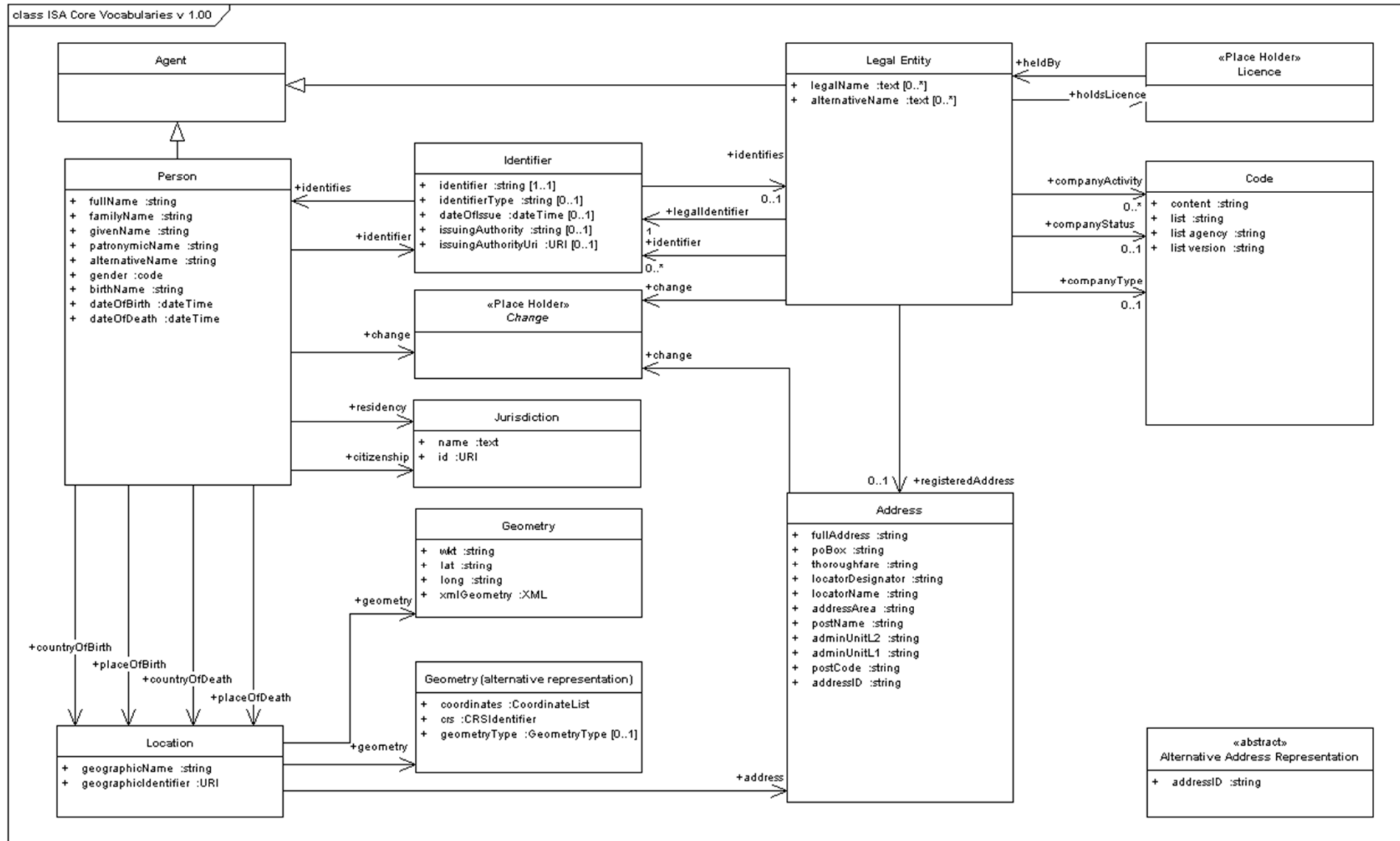
- 現在、初期検討が終わり、中核となるコアボキャブラリができたところ。
- 行政機関のメタデータスキーマを検討するADMSとも連携
- 電子政府コア・ボキャブラリは2012年後半に提案を作り2013年に向け更なる開発をしていく予定。

- 今後は、EIS、EIFとともにEIAの取り組みと融合していくものと考えられる。EIS、EIF、EIAの全体像は以下の構成になっており、全ての取り組みにセマンティック技術が重要な要素として取り入れられている。
 - European Interoperability Strategy (EIS)、European Interoperability Framework (EIF)、EUROPEAN INTEROPERABILITY ARCHITECTURE (EIA)、European Interoperability Infrastructure Services (EIS)
- EIA 2.0が公表された2011年11月には、コア・ボキャブラリが完成していなかったため簡単にしか記載されていないが、目指している方向は同じである。



- Action 1.1 (PROMOTING SEMANTIC INTEROPERABILITY AMONGST THE EUROPEAN UNION MEMBER STATES) に関連する分野で、 Open Source & Semantic Interoperabilityなどを扱うjoinupを、プラットフォームとして提供しており、ここには130のコミュニティが提案されている。
 - 活動状況は様々であり、分野もidからセマンティックスまでさまざまである。ベストプラクティスを収集し意見交換する仕組みであるePracticeとともに重要なプラットフォームとしての機能を提供している。
 - 2010Q3から2015Q4 でAction 1.1 に7millionEUROを投入。

Core Vocabularies Conceptual Model



ADMS (Asset Description Metadata Schema)

- デジタル資産のメタデータ体系も整理しており、検索や交換をしやすくしている。既存標準との連携を図ったり、企業と連携する等、注目される取り組みである。

Datasets extension	DCAT	ADMS	VoID
ds:DataCatalog	dcatalog:Catalog	adms:SemanticAssetRepository	
ds:DataDownload	dcatalog:Download	adms:SemanticAssetDistribution	
ds:Dataset	dcatalog:Dataset	adms:SemanticAsset	void:Dataset
ds:catalog		dcterms:isPartOf	
ds:dataset	dcatalog:dataset	dcterms:hasPart	
ds:distribution	dcatalog:distribution	radon:distribution	void:dataDump
ds:keyword	dcatalog:keyword	radon:keyword	
ds:license	dcterms:license	dcterms:license	
ds:spatial	dcterms:spatial	dcterms:spatial	
ds:temporal	dcterms:temporal		
sdo:about	dcatalog:theme	dcterms:subject	
sdo:contentSize	dcatalog:size		
sdo:contentUrl	dcatalog:accessURL	adms:accessURL	
sdo:copyrightHolder			
sdo:Country			
sdo:dateModified	dcterms:modified	dcterms:modified	
sdo:datePublished	dcterms:issued	dcterms:created	
sdo:description	dcterms:description	dcterms:description	
sdo:encodingFormat	dcterms:format	dcterms:format	
sdo:inLanguage	dcterms:language	dcterms:language	
sdo:name	dcterms:title	rdfs:label	
sdo:Organization	foaf:Organization		
sdo:Person	foaf:Person		
...

全体整理

NIEMとISAの比較

	ISA	NIEM
背景	行政機関のインタオペラビリティの確保が重要であった。	対テロ情報等を組織横断で情報共有する必要があった。
発足経緯	EUのインタオペラビリティのアーキテクチャの中で推進	司法省のxmlをベースに整備
実施体制	沢山	沢山
予算	26.8mユーロ	非公開
フレーム		
ボキャブラリ		
メタデータ	ADMS	
メッセージ	×	IEPD
API	×	×
バージョン	1	2（3に向けて作業中）
実装	FoaF等、本PJに取り込んでいる既存標準がある	
オープンデータとの距離	もともとインタオペラビリティからスタートしているので親和性は高い	テロ情報等のクローズなコアな部分から始まっているので、オープンデータとの融合はこれから

NIEMとISAのマッチング

- ISAのコアボキャブラリの多くはNIEMに含まれるが、一部対応しないものもある

NIEM AND ISA CORE VOCABULARY

ISA CORE VOCABULARY
-TO-
NIEM MAPPING

- 45 - Direct Matches
- 4 - No Match
- 2 - Partial Matches
- 3 - Uncertain

コンセプトの整理と日本への示唆

- まだ各国ともに模索している状況である。リソースが限られる中で最短距離で推進していく必要がある。

